Archiving Mars Mission Data Sets with the Planetary Data System

Report to MEPAG

Edward A. Guinness

Dept. of Earth and Planetary Sciences

Washington University

April 20, 2006

Outline

- Review lessons learned in archiving Mars mission data sets
- Overview of archiving process
  - Planning for archives
  - Archiving during operations
- Lessons learned from interfacing with
  - Projects
  - Data producers
  - Data users

Introduction

- Planetary Data System (PDS) works with two interfaces in archiving Mars Mission data sets
  - Mars Projects and instrument teams
  - Data users
- For projects and instrument teams, PDS provides advice and services for designing and generating archives
- For data users, PDS provides services for data search and distribution
  - Lessons learned from data users feed back to the data producer

Overview of Archiving Process - Planning

- **Mars Exploration Program has established a Data Management Plan**
  - **Available at http://pds-geosciences.wustl.edu/missions/mep/mepdmp.pdf**
  - **Sets guidelines for archive planning, data products to be archived, data release policy, including release schedule**
  - **Mars Projects develop archive plans that are responsive to the DMP**
- **Planning planetary data archives – involves both project and team interfaces with PDS**
- **Project writes Archive Plan to define high-level archiving responsibilities**
  - **Include table of standard data products for each instrument team**
  - **Include schedule of deliveries to PDS**
  - **As specified in the MEP DMP, data release schedule is a 3-month data collection period following by 6 months for archive generation and validation**

Overview of Archiving Process - Planning

- Project and PDS form data archive working group
  - PDS assigns a lead node as point contact with project
  - Currently there are active DAWG's for Odyssey, MER, MRO, Phoenix, and MSL
  - DAWG's meet on a regular basis, usually by teleconference
- Teams and PDS nodes establish points of contact
- Teams and PDS nodes generate Interface Control Document (ICD)
  - Provides detailed archive responsibilities for team and PDS node, e.g., define method of data delivery, who assembles data products into archive volumes
  - ICD signed by team, PDS node, PDS management, and project management
- Goal is to complete the Archive Plan and ICDs by the project CDR

Overview of Archiving Process - Planning

- Teams write Data Product and Archive Volume SIS for each data set
  - Defines the structure and content of data products and archive volume
  - Defines the PDS label for each data product
- Teams generate sample data products with PDS labels
- PDS nodes conduct peer reviews of SIS and sample data products
- Teams finalize SISs
  - Teams revise SIS documents and labels as needed based on peer review
  - Signed by teams, PDS nodes, PDS management, and project management
- Test delivery process from team to PDS node
  - Validate data flow defined in Archive Plan and ICD
  - Demonstrate that system can scale up to typical operational delivery volume
  - For MRO, there are four archiving tests scheduled between May 06 and Feb 07
- SIS should be finalized and signed by launch; with archive testing during cruise to Mars in coordination with project operations testing

Overview of Archiving Process - Operations

- Teams deliver data to PDS nodes according to Archive Plan schedule
  - Deliver complete archive volumes; including manifest with checksums
  - Delivery occurs with enough lead time so that data are released on date given in Archive Plan
  - After first release, incremental deliveries of volume contents (only deliver new files and files that have changed)
  - Examples: Odyssey releases data every 3 months (Jan, Apr, Jul, Oct). MER releases 90 sols of data roughly every 3 months. First MRO release in Jun 07 with first 30 days of PSP data
- PDS nodes validate delivery
  - Validate manifest and checksums
  - Validate volume contents for PDS standards and SIS compliance
  - Report validation results to teams
- PDS nodes release data
  - Put archive volumes online on PDS node web site
- PDS catalog is updated

Lessons for Projects

- Have an archive person at the mission level (Project Scientist or designee) to work with the lead PDS node to coordinate work across teams such as
  - Completing ICDs and SIS documentation
  - Schedules
  - Delivery tests
  - Action items, etc
  - Aware of mission schedules and milestones relevant to archive work -- e.g. peer reviews should be done in time for software to be completed for an operations readiness test.
- If some entity besides the team is making the products (e.g. MIPL, SOC), get them involved early on
  - Include these facilities in ICDs

Lessons for Data Producers

- Assign a team archive representative with the knowledge, authority, and time to do the work
- Have an ICD with the PDS node that spells out the deliverables and who's going to do what
- Complete SIS documents and their reviews on schedule (typically by launch for Mars missions)
  - Feeds into design and development of data processing pipelines
  - Supports archiving tests
- PDS node should work closely with the data producer to help keep archiving process on schedule and to resolve issues if they arise

Lessons Learned from Data Users

- Data users have a variety of levels in terms of expertise and interest
  - Small research groups without a lot of technical resources for software development and computer maintenance
  - Investigators who want to use derived data from multiple data sets for modeling or synthesis types of studies
  - Investigators who need to work with raw data to apply additional calibrations or other corrections
- Raw data are not enough; consider what derived products the community wants
  - Some users also want calibrated data and derived data products.
- Data formats need to be readily imported into common software tools that the community uses
- Provide data in suitable formats and with enough documentation about the instrument, and its calibration such that data sets are useable for the long term after the mission is over